

(19)日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11)特許出願公表番号

特表平10-511002

(43)公表日 平成10年(1998)10月27日

(51)Int.Cl.⁶

C 12 N 15/09
C 12 Q 1/68

識別記号

Z N A

F I

C 12 N 15/00
C 12 Q 1/68

Z N A A
A

審査請求 有 予備審査請求 有 (全 46 頁)

(21)出願番号 特願平9-512103
(86) (22)出願日 平成8年(1996)9月12日
(85)翻訳文提出日 平成10年(1998)3月12日
(86)国際出願番号 PCT/US96/14638
(87)国際公開番号 WO97/10363
(87)国際公開日 平成9年(1997)3月20日
(31)優先権主張番号 08/527, 154
(32)優先日 1995年9月12日
(33)優先権主張国 米国(US)
(31)優先権主張番号 08/544, 861
(32)優先日 1995年10月18日
(33)優先権主張国 米国(US)

(71)出願人 ザ ジョーンズ ホプキンス ユニバーシティースクール オブ メディシン
アメリカ合衆国 21205 メリーランド州
パルチモア、ラットランド アベニュー
720
(72)発明者 キンズラー、ケネス、ダブリュ。
アメリカ合衆国 21015 メリーランド州
ベルエア、ハルカーク ウェイ 1403
(72)発明者 ヴォグルステイン、パート
アメリカ合衆国 21208 メリーランド州
パルチモア、ブレトン ウェイ 3700
(74)代理人 弁理士 平木 祐輔 (外1名)

最終頁に続く

(54)【発明の名称】 遺伝子発現の逐次分析法

(57)【要約】

転写産物の迅速な定量および定性分析法である、遺伝子発現の逐次分析(serial analysis of gene expression: SAGE)を提供する。発現された遺伝子に対応する、短い限られた配列のタグを単離し、分析する。短期間(例えば、数時間)で1,000以上の限られたタグの配列を決定することで、細胞または組織の機能に特徴的な遺伝子発現パターンが明らかになる。さらに、SAGEは新規な転写産物および遺伝子に対応する新規な配列タグを同定・単離するための遺伝子検索ツールとして有用である。

【特許請求の範囲】

1. 少なくとも2つの限定されたヌクレオチド配列のタグをもつ単離されたオリゴヌクレオチド組成物であって、少なくとも1つのタグが少なくとも1つの発現された遺伝子に対応している上記組成物。
2. 前記のオリゴヌクレオチドが約1～200のジタグからなる、請求項1に記載の組成物。
3. 前記のオリゴヌクレオチドが約8～20のジタグからなる、請求項2に記載の組成物。
4. 遺伝子発現を検出する方法であって、
相補的デオキシリボ核酸(cDNA)オリゴヌクレオチドを作製し、
第1cDNAオリゴヌクレオチドから限定されたヌクレオチド配列の第1タグを単離し、かつ第2cDNAオリゴヌクレオチドから限定されたヌクレオチド配列の第2タグを単離し、
第1タグを第1オリゴヌクレオチドリンカー(該第1オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションのための第1配列を含む)に結合し、かつ第2タグを第2オリゴヌクレオチドリンカー(該第2オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションのための第2配列を含む)に結合し、そして
該タグのヌクレオチド配列を決定する、
ことを含んでなり、その際該タグが発現された遺伝子に対応している上記方法。
。
5. 第1オリゴヌクレオチドリンカーに結合された第1タグを第2オリゴヌクレオチドリンカーに結合された第2タグに連結してジタグを形成することをさらに含んでなる、請求項4に記載の方法。
6. ジタグオリゴヌクレオチドを増幅することをさらに含んでなる、請求項5に記載の方法。
7. ジタグのコンカテマー(鎖状体)を作製することをさらに含んでなる、請求項5に記載の方法。
8. 前記のコンカテマーが約2～200のジタグからなる、請求項7に記載の方法

。

9. 前記のコンカテマーが約8~20のジタグからなる、請求項8に記載の方法。
10. 第1および第2オリゴヌクレオチドリンカーが同じヌクレオチド配列を有する、請求項4に記載の方法。
11. 第1および第2オリゴヌクレオチドリンカーが異なるヌクレオチド配列を有する、請求項4に記載の方法。
12. 第1および第2オリゴヌクレオチドリンカーが次の配列：

5'-TTTTACCACTTATTCAATTGGTCCTCTCGCACAGGGACATG -3'
5'- ATGGTCAATAAGTTAAGCCAGGAGAGCGTGTCCCT -5'

または

5'-TTTTGTAGACATTCTAGTATCTCGTCAAGTCGGAAGGGACATG -3'
5'- AACATCTGTAAGATCATAGAGCAGTTAGCCTTCCCT -5'

- (ここで、AはジデオキシAである)を有する、請求項11に記載の方法。
13. 前記のリンカーが認識部位から離れた部位での切断を可能にする第2制限エンドヌクレアーゼ認識部位を含む、請求項4に記載の方法。
 14. 第2制限エンドヌクレアーゼがIIS型エンドヌクレアーゼである、請求項13に記載の方法。
 15. IIS型エンドヌクレアーゼがBsmFIおよびFokIよりなる群から選択される、請求項14に記載の方法。
 16. ジタグが約12~60塩基対である、請求項5に記載の方法。
 17. ジタグが約18~22塩基対である、請求項16に記載の方法。
 18. ポリメラーゼ連鎖反応（PCR）により増幅する、請求項6に記載の方法。
 19. PCR用のプライマーが

5' -CCAGCTTATTCAATTGGTCC-3' および
5' -CTAGACATTCTAGTATCTCGT-3'

- よりなる群から選択される、請求項18に記載の方法。
20. 遺伝子発現を検出する方法であって、
cDNAサンプルを第1制限エンドヌクレアーゼで切断し、ここで該エンド

ヌクレアーゼはcDNAの5'または3'末端の限定された位置でcDNAを切断することにより限定された配列のタグを生成するものであり、

限定された5'または3'cDNAタグを単離し、

該タグの第1プールを、増幅プライマーをハイブリダイズさせるための第1配列をもつ第1オリゴヌクレオチドリンカーと連結し、かつ該タグの第2プールを、増幅プライマーをハイブリダイズさせるための第2配列をもつ第2オリゴヌクレオチドリンカーと連結し、

該タグを第2制限エンドヌクレアーゼで切断し、

該タグの2つのプールを連結してジタグを形成し、そして

該タグのヌクレオチド配列を決定する、

ことを含んでなり、その際該タグが発現された遺伝子からのmRNAに対応している上記方法。

21. ジタグを増幅することをさらに含んでなる、請求項20に記載の方法。
22. 第1制限エンドヌクレアーゼがcDNA中に少なくとも1つの認識部位を有する、請求項20に記載の方法。
23. 第1制限酵素が4塩基対の認識部位を有する、請求項22に記載の方法。
24. 制限エンドヌクレアーゼがNlaIIIである、請求項23に記載の方法。
25. cDNAが捕捉手段を含む、請求項20に記載の方法。
26. 捕捉手段が結合エレメントである、請求項25に記載の方法。
27. 結合エレメントがビオチンである、請求項26に記載の方法。
28. 第1および第2オリゴヌクレオチドリンカーが同じヌクレオチド配列を有する、請求項20に記載の方法。
29. 第1および第2オリゴヌクレオチドリンカーが異なるヌクレオチド配列を有する、請求項20に記載の方法。
30. 第1および第2オリゴヌクレオチドリンカーが次の配列：

5'-TTTACCAAGCTTATTCAATTGCTCCTCTCGCACAGGGACATG	-3'
3'- ATGGTCGAATAAGTTAAGCCAGGAGAGCGTGTCCCT	-5'

または

5'-TTTTGTAGACATTCTAGTATCTCGTCAAGTCGGAAGGGACATG -3'
 3'-AACATCTGTAAGATCATAGAGCAGTCAGCCTTCCCT -5',

(ここで、AはジデオキシAである)を有する、請求項29に記載の方法。

31. 第2制限エンドヌクレアーゼがその認識部位から離れた部位で切断する、請求項20に記載の方法。

32. 第2制限エンドヌクレアーゼがIIS型エンドヌクレアーゼである、請求項31に記載の方法。

33. IIS型エンドヌクレアーゼがBsmFIおよびFokIよりなる群から選択される、請求項32に記載の方法。

34. ジタグが約12~60塩基対である、請求項20に記載の方法。

35. ジタグが約14~22塩基対である、請求項34に記載の方法。

36. ジタグを連結してコンカテマーを作製することをさらに含んでなる、請求項20に記載の方法。

37. 前記のコンカテマーが約2~200のジタグからなる、請求項36に記載の方法

。

38. 前記のコンカテマーが約8~20のジタグからなる、請求項37に記載の方法。

39. ポリメラーゼ連鎖反応(PCR)により増幅する、請求項20に記載の方法。

40. PCR用のプライマーが

5'-CCAGCTTATTCAATTCCGCC-3' および

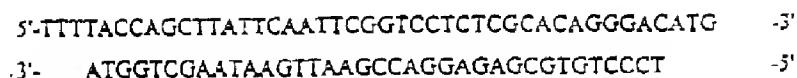
5'-GTAGACATTCTAGTATCTCGT-3'

よりなる群から選択される、請求項39に記載の方法。

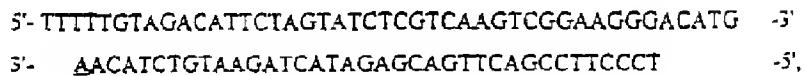
41. cDNAジタグの存在が該ジタグのタグの配列を有する遺伝子の発現を示すことからなる、遺伝子発現を検出するためのキットであって、増幅プライマーをハイブリダイズさせるための第1配列をもつ第1オリゴヌクレオチドリンクーを含有する第1容器、増幅プライマーをハイブリダイズさせるための第2配列をもつ第2オリゴヌクレオチドリンクーを含有する第2容器(これらのリンクーは制限エンドヌクレアーゼ認識部位から離れた部位でDNAを切断するための制限エンドヌクレアーゼ認識部位をさらに含む)、および該リンクーのユニークな第1および第2配列にハイブリダイズさせるための核酸プライマーを含有する第3お

および第4容器を含んでなる、1以上の容器から構成される上記キット。

42. 前記のリンカーが次の配列：



または



(ここで、AはジデオキシAである)を有する、請求項41に記載のキット。

43. 制限エンドヌクレアーゼがIIS型エンドヌクレアーゼである、請求項41に記載のキット。

44. IIS型エンドヌクレアーゼがBsmFIである、請求項43に記載のキット。

45. 増幅用プライマーが

5'-CCAGCTTATTCAATTCCGGTCC-3' および

5'-GTAGACATTCTAGTATCTCGT-3'

よりなる群から選択される、請求項41に記載のキット。

【発明の詳細な説明】

遺伝子発現の逐次分析法

本発明は、米国立衛生研究所の認可番号 CA57345、CA35494 およびGM07309 からの援助により行われたものである。米国政府は本発明に対していくらかの権利を有する。

本出願は、1995年9月12日付けの米国特許出願第08/527,154号の一部継続出願である。

発明の利用分野

本発明は、一般的には遺伝子発現の分野に関し、より特定すると、発現された遺伝子の領域に対応する転写産物の限定領域を同定することにより多数の転写産物を分析するための遺伝子発現の逐次分析 (serial analysis of gene expression: SAGE) 法に関する。

発明の背景

ヒトを含めた高等生物のゲノム配列の解析は今や現実に到達できる目標となっている。しかし、この解析は遺伝子の複雑性の一つのレベルを表すにすぎない。順序づけられた、時宜を得た遺伝子の発現は、生物の定義および生物学にとって等しく重要な、もう一つの複雑性のレベルを表している。

ヒト・ゲノムプロジェクトの一部として、mRNAから逆転写される相補的DNA (cDNA) の配列を解析することの役割が討論されてきた。というのは、ゲノム配列決定の提唱者が、あらゆる組織、細胞型および発生段階で発現される mRNA を一つのこらず見つけ出すことは至難の技であると主張し、さらにインtron領域と遺伝子間領域（制御および調節配列を含む）からの多くの価値ある情報が cDNA 配列決定により失われるだろうと指摘してきたからである（ヒトゲノムの地図作成および配列決定に関する委員会報告 Report of the Committee on Mapping and Sequencing the Human Genome, National Academy Press, Washington, D.C., 1988）。それゆえ、cDNA ライブライマーを用いるゲノムの

転写領域の配列決定では不十分であると考えられている。cDNA のライブルリーは反復性エレメント、ミトコンドリア遺伝子、リボソーム RNA 遺伝子、およ

び共通遺伝子やハウスキーピング遺伝子を含む他の核遺伝子が優勢であると考えられる。cDNAのライブラリーは構造および調節ポリペプチドまたはペプチドに対応する配列をすべて提供するとは限らないのである(Putneyら, *Nature*, 302: 718, 1983)。

標準的なcDNAクローニングのもう一つの欠点は、一部のmRNAが豊富であるのに対し、他が稀であるということである。さまざまな遺伝子に由来するmRNAの細胞量は数桁分も変化することがある。

cDNAサブトラクションまたはディファレンシャルディスプレーに基づいた技術は、2つの細胞型間の遺伝子発現を比較する上で極めて有用である(Hedrickら, *Nature*, 308:149, 1984; Liang and Pardee, *Science*, 257:967, 1992)。しかし、この技術は部分的な分析を提供するにすぎず、メッセンジャーRNAの量に関する直接的な情報をなにも提供しない。エクスプレスド・シークエンス・タグ(expressed sequence tag: EST)法は遺伝子を検索するための有用なツールであることがわかった(Adamsら, *Science* 252:1656, 1991; Adamsら, *Nature*, 355: 632, 1992; Okuboら, *Nature Genetics*, 2:173, 1992)ものの、ノーザンプロットティング、RNアーゼプロテクション、および逆転写酵素-ポリメラーゼ連鎖反応(RT-PCR)分析(Alwineら, *Proc. Natl. Acad. Sci. U.S.A.*, 74:5550, 1977; Zinnら, *Cell*, 34:865, 1983; Veresら, *Science*, 237:415, 1987)と同様に、一度に限られた数の遺伝子しか評価できない。さらに、EST法は、好ましくは、類似性検索やマッピングのために150塩基対またはそれより長いヌクレオチド配列を使用する。

また、配列標識部位(sequence tagged site: STS)(Olsonら, *Science*, 245:1434, 1989)は、ゲノムの物理的 地図を作成するためのゲノムマークターを同定するのに利用してきた。物理的にマップされたクローンからのこうした短い配列がゲノムにおける唯一同定された地図位置を表す。対照的に、発現された遺伝子の同定は、*in vivo*で実際に転写され発現された遺伝子のマークターである、発現された配列のタグをあてにしている。

さまざまな生物学的応用を研究するために、特に、いろいろな生理的または病

理的状態下にある異なる細胞型もしくは同じ細胞型における全体的な遺伝子発現パターンを確立するために、何千もの発現された遺伝子を迅速かつ詳細に分析することを可能とする改良法が必要とされている。異なる発現パターンの同定は、適切な治療ターゲットの同定、遺伝子治療（例えば、遺伝子置換）用の候補遺伝子の同定、組織タイピング、法的確認、疾病関連遺伝子の位置決定、診断・予診用のインジケーター遺伝子の同定を含めて、いくつかの実用性を有している。

発明の概要

本発明は、さまざまな生理的状態、発生段階または疾病状態にある異なる細胞型もしくは同じ細胞型における全体的な遺伝子発現パターンを同定するために、多数の転写産物を迅速に分析する方法を提供する。この方法はメッセンジャーRNAの限定された位置にある短いヌクレオチド配列のタグを同定することに基づいている。このタグを用いて、対応する転写産物および遺伝子（該遺伝子から該転写産物が転写される）を同定する。「ジタグ」(ditag)と称する二量体化されたタグを利用することで、本発明の方法は、クローニングおよび／または增幅中に生じ、おそらくはデータの評価中にも生じうる、いくつかのタイプの偏り(bias)を排除することができる。これらの短いヌクレオチド配列のタグを鎖状に連結することで、単一のDNA分子（例えば、ベクターまたは單一クローンに挿入されたDNA分子）上の複数のタグの配列決定を行うことにより逐次的に転写産物を効率よく分析することが可能である。

本明細書に記載する方法は、多数の転写産物の分析を可能とする新規な方法としての遺伝子発現の逐次分析(SAGE)法である。この戦略を検証するために、膵臓から単離したmRNAから短いcDNA配列タグを作製し、ランダムに対合させてジタグを形成し、鎖状に連結（コンカテマー化）してクローン化した。1000個のタグの手動配列決定により、膵臓の機能に特徴的な遺伝子発現パターンが明らかになった。このようなパターンの同定は例えば診断上および治療上重要である。さらに、遺伝子検索用ツールとしてのSAGEの使用を、新規なタグに対応する膵臓の新たな転写産物を同定・単離することで証明した。SAGEはさまざまな正常状態、

発生段階および疾病状態における発現遺伝子の定量的目録作成(cataloging)および比較のための広範に応用可能な手段を提供する。

図面の簡単な説明

図1は、SAGEの模式図を示す。この例において、第1制限酵素すなわちアンカリング酵素(anchoring enzyme)はNlaIIIであり、第2制限酵素すなわちタギング酵素(tagging enzyme)はFokIである。配列はプライマー由来の配列および転写産物由来の配列を表し、ここで「X」および「O」は異なるタグのヌクレオチドを表す。

図2は、転写産物の存在量の比較を示す。棒線はSAGE(黒線)またはハイブリダイゼーション分析(白線)で測定した存在量パーセント(percent abundance)を表す。SAGE数量化は表1から次のように誘導された。TRY1/2はトリプシノーゲン1および2のタグを含み、PROCARはプロカルポキシペプチダーゼA1のタグを示し、CHYMOはキモトリプシノーゲンのタグを示し、そしてELA/PROはエラスターEIIIBおよびプロテアーゼEのタグを含む。誤差線は、カウントした事象の平方根をとり、それを存在量パーセントに換算する(想定上のポアソン分布)ことで決定した標準偏差を表す。

図3は、SAGEタグによりcDNAライブラリーをスクリーニングした結果を示す。P2は実施例に記載したような13 bpのオリゴヌクレオチドを用いて得られた典型的なハイブリダイゼーション結果を示す。P1およびP2は表2に記載した転写産物に対応する。画像はMolecular Dynamics Phosphor Imagerを使って得られ、円はハイブリダイゼーションに先立って組換え体ファージを移行させたフィルターメンプランのアウトラインを示す。

図4は、本発明によるタグコード・データベース・アクセスシステムの構成図である。

好適な実施態様の説明

本発明は、発現された遺伝子に対応する転写産物の量および性質を決定するための迅速で定量的な方法を提供する。遺伝子発現の逐次分析(SAGE)と名づけられ

たこの方法は、遺伝子セグメントに対応する転写産物の限定された部分配列の同

定および特性決定を土台とするものである。これらの限定された転写産物配列「タグ」は、例えば細胞、組織または抽出物において発現される遺伝子のマーカーに相当する。

SAGEはいくつかの原理に基づいている。まず第一に、短いヌクレオチド配列のタグ(9~10 bp)は、それが転写産物内の限定された位置から単離されるという条件で、その転写産物を唯一無二に同定するのに十分な情報量を含むものである。例えば、9 bp 程度に短い配列は、タグ部位でのランダムヌクレオチド分布を仮定すれば、262,144 の転写産物 (4^9) を識別することができ、一方推定値はヒト・ゲノムが約80,000~200,000 の転写産物をコードすることを示唆する(Fieldら, *Nature Genetics*, 7:345, 1994)。下等な真核生物や原核生物の場合、例えばゲノムによりコードされる転写産物の数がより少ない場合は、タグのサイズをもっと短くすることができる。例えば、酵母では転写産物を識別するのに6~7 bp ほどの短いタグで十分である。

第二に、タグのランダム二量体化により偏り（增幅および／またはクローニングにより引き起こされる）を減らすことができる。第三に、これら短い配列のタグを鎖状に連結（コンカテマー化）することで、单一のベクターまたはクローン内の複数のタグの配列決定を行うことにより逐次的に転写産物を効率よく分析することができる。情報がデータの連続したつながりとして伝達されるコンピュータによる逐次コミュニケーションと同様に、配列タグの逐次分析は各タグの記録および境界を確立するための手段を必要とする。これらの原理はすべてを独立に使用しても、組み合わせて使用しても、他の公知の配列同定法と組み合わせて使用してもよい。

第一の実施態様において、本発明は、例えば特定の発生段階または特定の疾病状態での、特定の細胞もしくは組織または細胞抽出物における遺伝子発現を検出する方法を提供する。この方法は、相補的デオキシリボ核酸(cDNA)オリゴヌクレオチドを作製し、第1cDNAオリゴヌクレオチドから限定されたヌクレオチド配列の第1タグを、そして第2cDNAオリゴヌクレオチドから限定されたヌクレオチド配列の第2タグを単離し、第1タグを第1オリゴヌクレオチドリ

ンカー（該第1オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションのための第1配列を含む）に結合し、第2タグを第2オリゴヌクレオチドリンカー（該第2オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションのための第2配列を含む）に結合し、そして1以上の該タグのヌクレオチド配列を決定することを含んでなり、ここで該タグは発現された遺伝子に対応する。

図1は、本発明の方法において記載したSAGEを用いてメッセンジャーRNA（mRNA）を分析することを示した模式図である。mRNAの逆転写による二本鎖DNA配列のin vitro合成のために、対象となる細胞または組織からmRNAを単離する。mRNAの作製された二本鎖DNA相補体を相補的DNA（cDNA）という。

本明細書中で用いる「オリゴヌクレオチド」とは、2以上（好ましくは3以上）のデオキシリボヌクレオチドまたはリボヌクレオチドからなるプライマーまたはオリゴマー断片のことである。その正確なサイズは多くの要因に左右され、こうした要因もまたオリゴヌクレオチドの最終的な機能または用途次第で変化する。

本方法はさらに、第1オリゴヌクレオチドリンカーに結合された第1タグを、第2オリゴヌクレオチドリンカーに結合された第2タグに結合して、「ジタグ」を形成することを含んでなる。それぞれのジタグは少なくとも1つの遺伝子を代表する少なくとも1つの転写産物の2つの限定されたヌクレオチド配列に相当する。典型的には、ジタグは2つの異なる遺伝子からの2つの転写産物を表す。ジタグ内の限定されたcDNAタグの存在は、そのタグの配列を有する遺伝子の発現を示すこととなる。

増幅工程に先立って形成されたジタグを分析することにより、増幅（例えば、PCR）により導入される可能性があるひずみ(distortion)を排除することができる。ジタグの形成のためにタグを対合させることはランダムな事象である。異なるタグの数は多いと予想されるので、2つのタグが同一のジタグに結合される確率は豊富な転写産物のときでさえ低い。それゆえ、偏りのある(biased)標準的な増幅および／またはクローニング法により生成される可能性のある反復ジタグ

が本発明の方法による分析からは除外される。

「限定された」ヌクレオチド配列または「限定された」ヌクレオチド配列のタグなる用語は、転写産物の5'または3'末端から誘導されたヌクレオチド配列を指す。このヌクレオチド配列は第1制限エンドヌクレアーゼを用いて切断することにより限定され、どちらの末端が捕捉のために用いられるかに応じて、第1制限エンドヌクレアーゼ部位の5'または3'側のヌクレオチドを表す（例えば、本明細書に記載されるように、捕捉のためにオリゴ-dTを用いる場合は3'側のヌクレオチド）。

本明細書中で用いる「制限エンドヌクレアーゼ」および「制限酵素」なる用語は、認識部位または認識ヌクレオチド配列と称する特定の二本鎖DNA配列と結合し、その特定の認識部位でまたはその近傍で二本鎖DNAを切断する細菌の酵素を意味する。

「アンカリング酵素」または図1では「AE」と記される第1エンドヌクレアーゼは、転写産物を少なくとも1回切断して転写産物の5'または3'末端に由来する限定された配列のタグを生成する該酵素の能力により選択される。好ましくは、少なくとも1つの認識部位を有し、それゆえ大多数のcDNAを切断できる制限エンドヌクレアーゼが利用される。例えば、本明細書に例示するように、4塩基対の認識部位を有する酵素は平均して256塩基対（4⁴）ごとに切断すると予想されるが、大部分の転写産物はそれよりかなり大きい。4塩基対の部位を認識する制限エンドヌクレアーゼとしては、本発明の実施例に例示されるようなNlaIIIがある。DNA分子（例えば、cDNA）内に少なくとも1つの認識部位をもつ他の同様なエンドヌクレアーゼは当業者に公知である（例えば、Current Protocols in Molecular Biology, Vol. 2, 1995, Ausubelら編, Greene Publishing Association & Wiley Interscience, Unit 3.1.15; New England Biolabs Catalog, 1995を参照のこと）。

アンカリング酵素で切断した後、切断cDNAの最も5'または3'側の領域を、捕捉媒体に結合させることで単離することができる。例えば、本実施例に例示されるように、cDNA合成用のオリゴdTプライマーがビオチン化されている場合は、限定された3'ヌクレオチド配列タグをストレプトアビジンビーズを使って

単離する。この実施例では、第1酵素すなわちアンカリング酵素で切断すると、ポリA尾部の最も近くにある制限部位に対応する各転写産物上のユニーク部位が得られる。同様に、転写産物（そのcDNA）の5'キャップを、限定された5'ヌクレオチド配列タグを単離するための捕捉手段を標識または結合するために利用することができる。当業者であれば、ここに記載するような限定された配列のタグを単離するための他の同様な捕捉系（例えば、ビオチン／ストレプトアビシン、ジゴキシゲニン／抗ジゴキシゲニン）について熟知しているだろう。

本発明は单一の「アンカリング」エンドヌクレアーゼすなわち第1制限エンドヌクレアーゼの使用に制限されない。本発明の方法は、細胞または組織の完全な転写パターンを同定するために、別個の調製物サンプルに対して異なる酵素を用いて逐次的に実施することが望ましい。さらに、2以上のアンカリング酵素を用いると、第1アンカリング酵素から得られた発現パターンの確認が得られる。それゆえ、豊富な転写産物を表すcDNAがほとんどまたは全く切断されないように第1またはアンカリングエンドヌクレアーゼはcDNAをめったに切断しないことが想定される。かくして、切断される転写産物は「ユニーク」な転写産物に相当する。例えば、7～8bpの認識部位をもつ制限酵素はcDNAをめったに切断しない酵素であるだろう。同様に、完全な転写パターンを同定するために、以下に記載する2以上のタギング酵素を利用することができる。

本明細書中で用いる「単離された」なる用語には、ポリヌクレオチドが自然界でもともと結合している他の核酸、タンパク質、脂質、炭水化物または他の物質を実質的に含まないポリヌクレオチドが含まれる。cDNAはそのままで自然界に存在していないが、部分精製された天然に存在するmRNAを操作することで得られる。限定された配列のタグを単離するとは、切断された他のcDNAから5'または3'タグを精製することをいう。

一つの実施態様では、単離された限定ヌクレオチド配列タグを、リンカーの配列が異なるときは2つのcDNAプールに分ける。各プールをアンカリングすなわち第1制限エンドヌクレアーゼ部位で2つのリンカーのうちの1つに連結する。リンカーが同じ配列をもつときは、タグを2つのプールに分ける必要はない。第1オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションの

た

めの第1配列を含み、そして第2オリゴヌクレオチドリンカーは増幅プライマーのハイブリダイゼーションのための第2配列を含む。さらに、リンカーは「タギング酵素」または「TE」と記される第2制限エンドヌクレアーゼの認識部位を含む。本発明の方法は連結後にジタグオリゴヌクレオチドを増幅する必要がないが、増幅工程を含むことが好ましい。

第2制限エンドヌクレアーゼはその認識部位から離れた部位でまたは認識部位の外側で切断する。例えば、第2制限エンドヌクレアーゼはIIS型の制限酵素でありうる。IIS型制限エンドヌクレアーゼはその非対称認識部位から最大20 bp離れた一定の距離で切断する(Szybalski, W., Gene, 40:169, 1985)。IIS型制限エンドヌクレアーゼの例として、BsmFI およびFokIが挙げられる。その他の類似の酵素は当業者に公知である(Current Protocols in Molecular Biology, 前掲を参照のこと)。

限定されたヌクレオチド配列のタグに連結される第1および第2「リンカー」は同一のまたは異なるヌクレオチド配列をもつオリゴヌクレオチドである。例えば、本発明の実施例に示したリンカーには次のような異なる配列をもつリンカーが含まれる。

5'-TTTACCACTATTCAATTGGTCCTCTCGCACAGGGACATG -3'

(配列番号1)

3'- ATGGTCGAATAAGTTAAGCCAGGAGAGCGTGTCCCT -5'

(配列番号2)

および

5'-TTTGTAGACATCTAGTATCTCGCAAGTCGGAAGGGACATG -3'

(配列番号3)

3'- AACATCTGTAAGATCATAGAGCAGTTAGCCTTCCCT -5'

(配列番号4)、ここでAはジデオキシヌクレオチド(例えば、ジデオキシA)である。本発明の方法では他の類似のリンカーを用いてもよく、当業者であれば

、そのような代わりのリンカーを設計することができよう。

リンカーを設計するには、第2制限酵素すなわちタギング酵素による連結産物の切断により、限定されたヌクレオチド配列のタグ（例えば、本明細書に例示さ

れるような制限エンドヌクレアーゼ切断部位の3'側）をもつリンカーが放出されるようにする。限定されたヌクレオチド配列のタグは約6～30塩基対でありうる。好ましくは、タグは約9～11塩基対である。したがって、ジタグは約12～60塩基対となり、18～22塩基対とすることが好ましい。

同じ配列をもつリンカーに連結した限定タグの1つのプール、または異なるヌクレオチド配列をもつリンカーに連結した限定ヌクレオチド配列タグの2つのプールは互いとランダムに「尾一尾」連結させる。リンカーから最も遠いcDNAタグの部分を「尾」という。図1に示されるように、連結されたタグ対すなわちジタグは、ジタグの上流（5'側）に第1制限エンドヌクレアーゼ部位および下流（3'側）に第1制限エンドヌクレアーゼ部位、ジタグの上流と下流に第2制限エンドヌクレアーゼ切断部位、そしてジタグの上流と下流に第2制限酵素認識部位および増幅プライマーハイブリダイゼーション部位の両方を含むリンカーオリゴヌクレオチドをもっている。言い換えると、このジタグはそれぞれ第1制限エンドヌクレアーゼ部位、第2制限酵素認識部位およびリンカーにより挟まれている。

ジタグは各リンカーの一鎖に特異的にハイブリダイズするプライマーを利用して增幅することができる。好ましくは、米国特許第4,683,195号に記載されるような標準的なポリメラーゼ連鎖反応（PCR）法により増幅を行う。あるいはまた、原核生物に適合するベクターへのクローニングまたは当業者に公知の他の増幅法によりジタグを増幅してもよい。

本明細書中で用いる「プライマー」とは、核酸鎖に相補的なプライマー伸長産物の合成が誘導される条件（すなわち、ヌクレオチドおよびDNAポリメラーゼのような重合剤の存在下に適当な温度およびpHの条件）下に置かれたとき、合成の開始点として作用することができる、天然に存在するまたは合成的に製造されたオリゴヌクレオチドのことである。最大の増幅効率をあげるためにプライマ

ーは一本鎖で、しかもオリゴデオキシリボヌクレオチドであることが好ましい。

プライマーは重合剤の存在下で伸長産物の合成を開始させるに足る長さのものでなければならない。プライマーの正確な長さは温度およびプライマー源を含めた多くの要因により左右されるだろう。

ここで用いるプライマーは増幅しようとする特定配列のそれぞれの鎖に「実質

的に」相補的となるように選択される。このことは、プライマーがそれぞれの鎖とハイブリダイズするのに十分な相補性をもたねばならないことを意味する。それゆえ、プライマー配列は鋳型の正確な配列を反映する必要はない。本発明において、プライマーはオリゴヌクレオチドリンクーに対し実質的に相補的なものである。

配列番号1～4で示されるリンクーを増幅するのに有用なプライマーとしては、5'-CCAGCTTATTCAATTCCGTCC-3'（配列番号5）および5'-GTAGACATTCTAGTATCTCGT-3'（配列番号6）が挙げられる。当業者であれば、過度の実験を行わなくとも、リンクーのヌクレオチド配列に基づいて類似の増幅用プライマーを作製することができよう。

第1制限エンドヌクレアーゼで増幅PCR産物を切断することにより、鎖状に連結することでコンカテマー化しうるジタグを単離することができる。連結後、コンカテマーをクローニングすることが望ましい場合があるが、本発明の方法では必要でない。増幅を行おうと行うまいと、ジタグまたはコンカテマーは標準的な配列決定法により分析できる。コンカテマーは一般的に約2～200のジタグからなり、約8～20のジタグからなることが好ましい。これらは好適なコンカテマーであるが、コンカテマー化され得るジタグの数は個々のタグの長さに依存し、過度の実験を行わなくとも当業者であれば容易に決定できよう。コンカテマーの形成後、複数のタグを配列分析のためにベクターにクローニングすることができ、またジタグもしくはコンカテマーをクローニングを行わずに当業者に公知の方
法で直接配列決定にかけてもよい。

本発明の限定されたヌクレオチド配列のタグをクローニングする標準的な方法は、特に、プラスミドやファージのようなベクターにタグを挿入することである

。ここに記載の方法により作製されたジタグまたはジタグのコンカテマーは、配列分析、タグをプローブとして用いるラークノプラスミドハイブリダイゼーションといった後続の分析のために組換えベクター中に当業者に公知の方法でクローニングされる。

「組換えベクター」なる用語は、ジタグ遺伝子配列の挿入または組み込みにより操作された当技術分野で公知のプラスミド、ウイルスまたは他の運搬体を指す。

このようなベクターは例えばマーカー遺伝子配列の効率のよい転写を促進するプロモーター配列を含有する。典型的には、ベクターは複製起点、プロモーター、および形質転換細胞の表現型選択を可能とする特定の遺伝子を含有する。本発明で用いるのに適したベクターとして、例えば、pBlueScript (Stratagene, LaJolla, CA)、pBC、pSL301 (Invitrogen) および当業者に公知の他の同様なベクターがある。好ましくは、ジタグまたはそのコンカテマーは配列決定用のベクターに連結される。

ジタグをクローニングするためのベクターは適当な宿主細胞に移入できるものである。「宿主細胞」とは、その細胞内でベクターが増殖できかつそのDNAが発現され得る細胞のことである。この用語は宿主細胞自体の子孫をも意味する。複製の間に突然変異が起こることがあるので、すべての子孫が親細胞と同一であるとは限らないことが理解されよう。しかしながら、「宿主細胞」なる用語が用いられるとき、この用語にはこの種の子孫も含まれる。安定した移入法（外来DNAが宿主内に連続して維持されることを意味する）は当技術分野で公知である。

ジタグを含有するベクターによる宿主細胞の形質転換は、当業者には周知であるような、慣用の技法により行うことができる。宿主が大腸菌のような原核細胞である場合は、指数増殖期後に収穫し、続いて当技術分野で公知の手法によるCaCl₂法で処理した細胞から、DNA取込み能のあるコンピテント細胞を調製する。これとは別に、MgCl₂またはRbClを使用してもよい。また、エレクトロポレーションや当技術分野で常用される他の方法により形質転換を行うこともできる。

個々のクローン内に存在するジタグは、手動でまたは自動化法を使って、標準的な方法（例えば、*Current Protocols in Molecular Biology*, 前掲, Unit 7)により配列決定することができる。

もう一つの実施態様において、本発明は、限定されたヌクレオチドタグまたはジタグの存在が該タグの配列を有する遺伝子の発現を示すことからなる、遺伝子発現を検出するためのキットを提供し、このキットは、増幅プライマーをハイブリダイズさせるための第1配列をもつ第1オリゴヌクレオチドリンカーを含有する第1容器、増幅プライマーをハイブリダイズさせるための第2配列をもつ第2オリゴヌクレオチドリンカーを含有する第2容器（これらのリンカーは制限エン

ドヌクレアーゼ認識部位から離れた部位でDNAを切断するための制限エンドヌクレアーゼ認識部位をさらに含む）、および該リンカーのユニークな第1および第2配列にハイブリダイズさせるための核酸プライマーを含有する第3および第4容器を含んでなる1以上の容器から構成される。オリゴヌクレオチドリンカーが同じヌクレオチド配列をもつのであれば、本発明のキットではリンカーを含有する容器が1つだけ必要となる。

さらに別の実施態様において、本発明は、少なくとも2つの限定されたヌクレオチド配列のタグを有するオリゴヌクレオチド組成物であって、前記タグの少なくとも1つが少なくとも1つの発現された遺伝子に対応している前記組成物を提供する。この組成物は約1～200のジタグ、好ましくは約8～20のジタグからなる。このような組成物は、例えば細胞、組織または細胞抽出物中の発現遺伝子に対応する限定されたヌクレオチド配列のタグを同定することにより遺伝子発現を分析するのに有用である。

本発明のSAGE法を用いて示差的に発現された遺伝子を同定するにあたって、他のゲノム法と組み合わせて使用することが考えられる。例えば、個々のタグ（好ましくはジタグ）を固相支持体（例えば、ニトロセルロースフィルター、ガラススライド、シリコンチップ）上に固定したオリゴヌクレオチドとハイブリダイズさせることができる。こうした技法として、以下に記載するような「平行配列分析」(parallel sequence analysis)すなわちPSAがある。本発明の方法により

形成されたジタグの配列はまた、クローン配列決定（C S）を含めた方法により限界希釈を用いて決定することもできる。

簡単に述べると、P S Aはジタグの作製後に行われるが、その際ジタグがハイブリダイズするオリゴヌクレオチド配列は標識せずに、ジタグを検出可能に標識することが好ましい。あるいはまた、ジタグではなくオリゴヌクレオチドを標識してもよい。例えば、ラジオアイソトープ、蛍光化合物、生物発光化合物、化学発光化合物、金属キレート剤、または酵素によりジタグを検出可能に標識することができる。当業者であれば、ジタグに結合させるのに適した他の標識について熟知しており、ルーチンな実験操作を用いてそれを確かめることができよう。例えば、標識した（例えば、フルオレセインで標識した）プライマーを用いてPCRを行ってもよい。ジタグは蛍光末端標識をもつことが好ましい。

標識したまたは未標識のジタグを一本鎖分子に分離し、それを好ましくは段階的に希釈し、そして、例えば10-merのあらゆる可能な順列を表すオリゴヌクレオチドを（例えば、シリコンチップの各グリッド中に）含有する固相支持体（例えば、Fodorら, Science, 251:767, 1991に記載されるシリコンチップ）に添加する。次に、固相支持体上のオリゴヌクレオチドを、異なる状態にある細胞（例えば、異なる発生段階、増殖因子の不在下および存在下での細胞の増殖、正常細胞対形質転換細胞、さまざまな組織発現の比較など）から得られたタグとハイブリダイズさせることにより、該支持体内（例えば、チップのグリッド上）に含まれるタグの示差的発現を決定する。フルオレセインで末端標識したタグの場合は、蛍光の分析が特定の10-merへのハイブリダイゼーションを示す。例えば、固定したオリゴヌクレオチドがフルオレセインで標識されている場合は、（標識オリゴにハイブリダイズしたジタグの接近による）消光ゆえの蛍光の減少が観察され、遺伝子発現のパターンに関して分析される。この方法の代表例を本明細書中の実施例4に示す。

本発明のSAGE法は、細胞系のクローニングに用いられる限界希釈法と類似したクローン配列決定にも有用である。例えば、ジタグまたはそのコンカテマーを希釈して、各受け器が1未満のDNA分子を含むように各受け器に入れる。各受け

器中のDNAを増幅し、マススペクトロスコピーを含めた当技術分野で公知の標準方法により配列決定する。示差的発現の評価はSAGEに関して上述したようを行う。

当業者であれば、過度の実験によらなくとも、本発明において記載したSAGEにより得られたジタグまたは個々のタグを分析するための他の方法を簡単に決定できるだろう。

本発明にしたがってある配列から限定されたタグを導き出す概念は、配列データベースにサンプルのタグを一致させるのに有用である。好適な実施態様では、サンプル配列と既知配列とを一致させるためにコンピュータ法が用いられる。

一つの実施態様では、サンプルの配列タグを、配列データベース中の対応する情報と比較して、サンプル配列と一致する既知配列を同定する。配列データベ

ース中の各配列につき 1 以上のタグを該配列内の各アンカリング酵素部位に隣接する N 個の塩基対として決定することができる。しかし、好適な実施態様では、3' 末端から最初のアンカリング酵素部位のみを用いてタグを決定する。好ましくは、タグを限定する隣接塩基対はアンカリング酵素部位の 3' 側にあり、かつ N は 9 である。

このようなデータベースに対しての線形検索(Linear search)を用いることができる。しかし、好適な実施態様では、N 塩基タグの各塩基対 (A、C、G または T) を数字または「タグコード」(例えば、A = 0、C = 1、G = 2、T = 3、または他の適当な符号) に変換することで、サンプルからの配列タグをユニークな数値表示に変える。上記のような配列データベースの各配列につき 1 つのタグを決定し、そのタグを同様な方法でタグコードに変換する。好適な実施態様では、配列データベースのタグコードのセットをポインターファイル(pointer file)に保存する。サンプル配列のタグコードをポインターファイル中のタグコードと比較して、サンプルタグコードに対応する配列の配列データベース中の位置を決定する。(配列データベースが冗長性(redundancy)をもつ場合は複数の対応配列が存在しうる。)

図 4 は、本発明によるタグコード・データベース・アクセスシステムの構成図

である。配列データベース10（例えば、ヒト・ゲノム配列データベース）を上記のように処理して、各配列が決定されてポインターファイル12に保存されたタグコードをもつようとする。サンプルのタグコードXを上記のように決定し、コンピュータの記憶位置14に保存する。一致する配列タグコードに関してサンプルのタグコードXとポインターファイル12とを比較する。一致が見つかれば、その一致する配列タグコードと関連したポインターを用いて、配列データベース10中の対応配列にアクセスする。

ポインターファイル12はいくつかのフォーマットのいずれであってもよい。一つのフォーマットでは、ポインターファイル12の各エントリーがタグコードと、配列データベース10中の対応する記録に対するポインターを含む。サンプルのタグコードXは配列タグコードと線形検索で比較することができる。あるいはまた、配列タグコードを分類して、バイナリー検索(binary search)を用いてもよい。

もう一つの代替フォーマットとして、分類体系的な樹木構造（例えば、B樹木）に、または単一にもしくは二重にリンクされたリストとして、あるいは他の簡便に検索できるデータ構造もしくはフォーマットで、配列タグコードを構造化することもできる。

好ましい実施態様において、ポインターファイル12の各エントリーは配列データベース10中の対応する記録に対するポインターだけを含む。ポインターファイル12を構築するにあたって、タグコードの数値に対応するポインターファイル12中のエントリー位置に各配列タグコードを割り当てる。例えば、配列タグコードが「1043」であったとすると、配列データベース10中の対応する記録に対するポインターはポインターファイル12のエントリー#1043に保存されるだろう。サンプルのタグコードXの数値を用いて、サンプルタグコードXに対応するポインターファイル12中の該位置を直接アドレス指定し、ひいては配列データベース10をアドレス指定するために該位置に保存されたポインターに迅速にアクセスすることができる。

すべての可能な塩基対を表すのにたった4つの数値しか必要でないので、好適

なポインターファイル12構造体と共にタグコードのバイナリー・コーディド・デシマル(binary coded decimal: BCD)数を用いることは、記憶または蓄積スペースをむだにする「希薄」なポインターファイル12をもたらす。したがって、本発明は公知の様式で各タグコードを基底4(すなわち、コードディジットにつき2ビット)に変換して、結果的にコンパクトなポインターファイル12構造体が得られる。例えば、タグ配列「A G C T」(ただし、A=00₂、C=01₂、G=10₂、T=11₂)に関して、バイナリーでの基底4表示は「00011011」となるだろう。これに対して、BCD表示は「00000000 00000001 00000010 00000011」となるだろう。もちろん、塩基対の他の符号化も同様の機能を与えることを理解すべきである。

本発明にしたがってサンプル配列から限定されたタグを導き出す概念は、さまざまなサンプルを類似性に関して比較するにも有用である。好ましい実施態様では、異なるサンプルからの配列タグを一致させるためにコンピュータ法を用いる。例えば、多数の配列を含む材料(例えば、組織)を比較するにあたって、第1サ

ンプル中の種々のタグの存在頻度を、分布またはヒストグラム型のデータ構造中に保存されたタグコードとしてしるすことができる。例えば、図4のポインターファイル12と同様に構造化された表を使用でき、この表では各エントリーが存在頻度の値を含む。その後、第2サンプル中の種々のタグを作製し、タグコードに変換し、そのタグコードにより表エントリーを直接アドレス指定することで表と比較することができる。出力装置にテキストまたはグラフの形で情報を出力するために、および/または、あとで使用するべくデータ保存システムに保存するために、一致した数のみならず一致した位置を計測していく。

本発明のタグ比較の面はハードウェアもしくはソフトウェア、または両者の組合せで実行することができる。好ましくは、本発明のこれらの面は、プロセッサ、データ保存システム(揮発性および非揮発性の記憶および/または蓄積素子を含む)、少なくとも1つの入力装置、および少なくとも1つの出力装置から構成されるプログラム可能なコンピュータで作成するコンピュータプログラムにより

実行される。データ保存システムへの一時的または永久的な保存のための1以上の入力装置からのデータ入力は配列を含み、また既知および／または未知配列の以前に作製されたタグおよびタグコードを含みうる。入力データにプログラムコードを当てはめて、上記の機能を遂行しつつ出力情報を引き出す。この出力情報を公知の様式で1以上の出力装置に適用する。

このようなコンピュータプログラムはそれを、蓄積媒体または装置をコンピュータで読み取ってここに記載の手順を実行するときそのコンピュータをコンフィギュア(configuring)して操作するために、一般用または専門用のプログラム可能なコンピュータで読み取り可能な蓄積媒体または装置（例えば、ROMまたはフロッピーディスク）に保存することが好ましい。本発明のシステムはまた、コンピュータプログラムを伴った、コンピュータで読み取り可能な蓄積媒体として実施することも考えられ、その場合は、このような蓄積媒体により、上記の機能を果たすように予め決められた特別なやり方でコンピュータを操作する。

以下の実施例は本発明を例示するもので限定するものではない。それらは用いることのできる手順を代表するが、当業者に公知の他の手順も代わりに使用できるだろう。

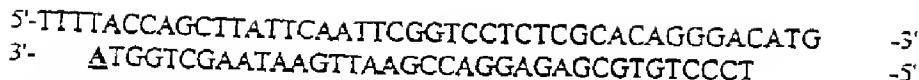
実施例

例示目的のために、本発明のSAGE法を用いてヒト臍臍における遺伝子発現を特性づけた。第1制限エンドヌクレアーゼすなわちアンカリング酵素としてNlaIIIを、第2制限エンドヌクレアーゼすなわちタギング酵素としてBsmFIを使用して9 bp のタグを得た。BsmFIは認識部位 GGGACの14 bp 3'側の相補鎖を切断し、4 bp の5'突出部分を生成させると予測された(New England BioLabs)。示した(GGGACATG)ようにBsmFI部位とNlaIII (CATG) 部位を重ね合わせると11 bp のタグが得られると予測される。しかし、用いた切断条件 (37°C) 下で、BsmFIはしばしばその認識部位のより近くで切断して、その認識部位の3'側の最小12 bp を残すことが分析から示唆された。それゆえ、タグの分析のためにアンカリング酵素部位に最も近い9 bp だけを使用した。65°Cで切断すると、より一貫して11bp のタグが得られる。

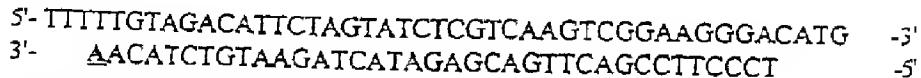
Gen Bankから得られたヒト転写産物のコンピュータ分析は、9 bp の長さのタグの95%以上がユニークであるらしく、追加の2個の塩基を含めても更なる分解能がほとんど得られないことがわかった。IntelliGenetics Bionetオンラインサービスで提供されたFindseqプログラムを使って、GenBank 87データベースからヒト配列(84,300)を抽出した。以後の分析はすべてMicrosoft Windows 操作システム用のMicrosoft Visual Basicに書かれたSAGEプログラムグループを用いて行った。SAGEデータベース分析プログラムは位置記載(locus description)中に「RNA」として示した配列のみを含め、「EST」として示したエントリーを除くようにセットし、その結果13,241の配列に減少した。アンカリング酵素としてNlaIIIを用いてこのサブセットの配列を分析したところ、4,127 の9 bp タグがユニークである一方で、1,511 のタグは1より多いエントリーで見いだせることができた。後者のエントリーの無作為選択したサブセット(100)のスクレオチド比較から、少なくとも83%が同一の遺伝子または高度に関連した遺伝子(少なくとも250 bpにわたって>95%の同一性)の冗長データベースのエントリーによることが示された。これにより、9 bp タグのうち5381 (95.5%) が転写産物または高度に保存された転写産物ファミリーに対しユニークであると示唆された。

実施例 1

上記で概説したように、ヒト臍臓由来のmRNAを用いてジタグを作製した。簡単に説明すると、5 μgの全臍臓由来のmRNA(Clontech)を、BRL cDNA合成キットを用いて、製造業者のプロトコールに従い、プライマーとしてビオチン-5' T₁₈-3' を用いて二本鎖cDNAに変換した。次に、このcDNAをNlaIIIで切断し、磁気ストレプトアビジン・ビーズ(Dynal)に結合することにより3' 制限断片を単離した。結合したDNAを2つのプールに分け、以下のリンクーの中の1つを各プールに連結した。



(配列番号1および2)



(配列番号3および4)

(ここで、Aはジデオキシヌクレオチド（例えば、ジデオキシA）である。)

丹念に洗浄して未連結のリンカーを除去した後、BsmFIで切断することによってこれらのリンカーおよび隣接するタグを遊離させた。生成した突出部分をT4ポリメラーゼでフィルインし、これらのプールを合わせて互いに連結した。次に、目的の連結産物を、プライマーとして

5'-CCAGCTTATICAATTGGTCC-3'

および

5'-GTAGACATTCTAGTATCTCGT-3'

(それぞれ配列番号5および6)を用いて25サイクルにわたって増幅した。次に、このPCR反応をポリアクリルアミドゲル電気泳動で分析し、目的の産物を切り出した。次に、さらに15サイクルのPCRを行い、効率的な連結およびクローニングに十分な量の産物を得た。

PCRジタグ産物をNlaIIIで切断し、これらジタグを含有するバンドを切り出し、自己連結させた。連結後、ポリアクリルアミドゲル電気泳動によってコン

カテマー形成ジタグを分離し、200bpより大きな産物を切り出した。これらの産物をpSL301(Invitrogen)のSphI部位にクローニングした。クローニング部位の外側のT7およびT3配列をプライマーとして用いるPCRにより、コロニーをインサートについてスクリーニングした。少なくとも10(10~50の範囲)のタグを含有するクローンを、記載(Del Salら、Biotechniques 7:514、1989)のようにして、5'-GACGTCGACCTGAGGTAAATTATAACC-3'（配列番号7）をプライマーとして用いて、PCR増幅によって同定し、手操作により配列決定した。SAGEソフトウェア・グループを用いて配列のファイルを分析した。このソフトウェア・グル

ープは、適切なスペーシングを有するアンカリング酵素部位を同定し、2つの介在タグを抽出し、それらをデータベースに記録するものである。413のユニークなジタグおよび87の反復ジタグから1,000のタグを誘導した。後者（反復ジタグ）は1回だけ計測することで、定量で起こりうるPCRの偏りを排除した。SAGEソフトウェアの機能は、単に、遺伝子配列の検索を最適化することだけである。

表1は、最初の1,000タグの分析を示す。16%が配列のアンピギュイティー(ambiguity)を有していたか、あるいはリンカー配列から誘導されたものであるので、それらを除外した。残りの840のタグは、一回存在した351のタグおよび複数回見出された77のタグを含んでいた。最も豊富なタグ10のうち9がGenBankR87中の少なくとも1つのエントリーに一致した。その後、残りのタグはアミラーゼに由来することがわかった。10の転写産物の全部が既知の臍臓機能の遺伝子に由来し、それらの存在は、慣用のアプローチを用いたこれまでの臍臓RNAの分析と一致した(Hanら、Proc.Natl.Acad.Sci.U.S.A.83,:110,1986; Takedaら、Hum. Mol. Gen. 2:1793, 1993)。

表1
脾臓 SAGE タグ

<u>タグ</u>	<u>遺伝子</u>	<u>N</u>	<u>パーセント</u>
GAGCACACC	プロカルボキシペプチダーゼ A1(X67318)	64	7.6
TTCTGTGTG	脾臓トリプシノーゲン 2(M27602)	46	5.5
GAACACAAA	キモトリプシノーゲン(M24400)	37	4.4
TCAGGGTGA	脾臓トリプシン 1(M22612)	31	3.7
GCGTGACCA	エラスターーゼ IIIB(M18692)	20	2.4
GTGTGTGCT	プロテアーゼ E(D00306)	16	1.9
TCATTGGCC	脾臓リバーゼ(M93285)	16	1.9
CCAGAGAGT	プロカルボキシペプチダーゼ B(M81057)	14	1.7
TCCTCAAAA	一致なし、表2、P1 参照	14	1.7
AGCCTTGGT	胆汁塩刺激リバーゼ(X54457)	12	1.4
GTGTGCGCT	一致なし	11	1.3
TGCGAGACC	一致なし、表2、P2 参照	9	1.1
GTGAAACCC	21 Alu エントリー	8	1.0
GGTGACTCT	一致なし	8	1.0
AAGGTAACA	分泌トリプシンインヒビター(M11949)	6	0.7
TCCCCCTGTG	一致なし	5	0.6
GTGACCACG	一致なし	5	0.6
CCTGTAATC	M91159, M29366, 11 Alu エントリー	5	0.6
CACGTTGGA	一致なし	5	0.6
AGCCCTACA	一致なし	5	0.6
AGCACCTCC	伸長因子 2 (Z11692)	5	0.6
ACGCAGGGA	一致なし、表2、P3 参照	5	0.6
AATTGAAGA	一致なし、表2、P4 参照	5	0.6
TTCTGTGGG	一致なし	4	0.5

TTCATAACAC	一致なし	4	0.5
GTGGCAGGC	NF-κB(X61499), Alu エントリー(S94541)	4	0.5
GTAAAACCC	TNF 受容体 II (M55994)、 Alu エントリー(X01448)	4	0.5
GAACACACA	一致なし	4	0.5
CCTGGGAAG	臍臓ムチン(J05582)	4	0.5
CCCATCGTC	ミトコンドリア CytC オキシダーゼ(X15759)	4	0.5

(配列番号 8 ~ 37)

まとめ

S A G E タグ	3回より多い	380	45.2
存在	3回 (15 × 3 =)	45	5.4
	2回 (32 × 2 =)	64	7.6
	<u>1回</u>	<u>351</u>	<u>41.8</u>
総 S A G E タグ		840	100.0

「タグ」は、4 bpのアンカリングNlaIII部位に隣接する、各タグに固有の9 bpの配列を示す。「N」および「パーセント」は、それぞれ、タグを同定した回数およびその頻度を示す。「遺伝子」は、S A G E ソフトウェア・グループを用いて表示のタグに一致することがわかったGenBank R87エントリーの受け入れ番号および記載を示し、以下の例外がある。すなわち、重複するエントリーのために複数のエントリーが同定される場合には、1つのエントリーのみを掲載する。キモトリプシノーゲンおよびトリプシノーゲン1の場合、同じタグを含有すると予想される他の遺伝子が同定されたが、続いて行ったハイブリダイゼーションおよび配列分析によって、掲載の遺伝子をタグの供給源として同定した。「Aluエントリー(Alu entry)」とは、alu共通配列 (Deiningerら、J.Mol.Biol.151,:17,1981) の少なくとも1コピーを含有していた転写産物のためのGenBankエントリーと一致したことを示す。

実施例 2

S A G E の定量的性質は、オリゴ-dTプライマーを用いて作製 (oligo-dTprime

d) 腺臓 c DNA ライブライリー（トリプシノーゲン1/2、プロカルボキシペプチダーゼA1、キモトリプシノーゲンおよびエラスターーゼI～IIIB／プロテアーゼEに対するc DNA プローブでスクリーニングした）を構築することによって評価した。実施例1においてSAGEに用いたのと同じ調製物から得られた腺臓mRNA を用いて、ZAP Express c DNA合成キット（Stratagene）を用いて製造業者のプロトコールにしたがって、ZAP Expressベクターにc DNA ライブライリーを構築した。ランダムに選んだ15のクローンを分析することにより、100%がc DNA インサートを含有することがわかった。250～500のブラークを含有するプレートを、先に記載されている（Ruppertら、Mol. Cell. Biol. 8:3104,1988）ようにしてハイブリダイズさせた。トリプシノーゲン1、トリプシノーゲン2、プロカルボキシペプチダーゼA1、キモトリプシノーゲンおよびエラスターーゼIIIBに対するc DNA プローブを、RT-PCRによって腺臓RNAから誘導した。トリプシノーゲン1および2のプローブは93%同一であり、使用条件下で同じブラークにハイブリダイズした。同様に、エラスターーゼIIIBプローブとプロテアーゼEプローブは95%以上同一であり、同じブラークにハイブリダイズした。

これらの転写産物に対するSAGEタグの相対的存在量は、ライブライリー・スクリーニングを用いて得た結果と非常によく一致した（図2）。さらに、トリプシノーゲン1と2も、エラスターーゼIIIBとプロテアーゼEも、ライブライリーをスクリーニングするのに用いたc DNA プローブによって区別することはできなかつたが、4つの転写産物の全ては、それらのSAGEタグに基づいて容易に区別できた（表1）。

実施例3

SAGEは、既知の転写産物の存在量についての定量的情報を提供するため

だけでなく、新規な発現遺伝子の同定にも用いることができる。本実施例におけるSAGE分析の目的では、各転写産物に固有の9 bpの配列のみを考慮したが、各SAGEタグでは、アンカリング酵素(4 bp)部位および9 bpのタグから成る13 bpの配列が確定された。この可能性を説明するために、13bpのオリゴヌクレオチドを用いて、4つの指定を受けていないタグ（P1～P4）（すなわち、GenBankR87

からのエントリーに対応しないタグ)に対応する転写産物を単離した(表1)。それら4つの場合の各々において、13bpのオリゴヌクレオチドをハイブリダイゼーション・プローブとして用いて臍臓cDNAライブラリーをスクリーニングするだけで、タグに対する複数のcDNAクローンを単離することが可能であった(図3の例)。

250~2,000のplaquesを含有するプレートを、ハイブリダイゼーション温度を室温まで低下させた以外は標準的なプローブについて以前に記載されているものと同じ条件を用いて、オリゴヌクレオチド・プローブにハイブリダイズした。洗浄を6×SSC/0.1%SDS中で30分間室温で行った。これらのプローブは、T4ポリヌクレオチドキナーゼを用いて γ^{32} P-ATPで標識された13bpのオリゴヌクレオチドから構成された。それぞれの場合、誘導したクローンを配列決定することにより、同定した転写産物の想定3'末端に正しいSAGEタグが同定された。13-merとのハイブリダイゼーションによって同定されたplaquesの存在量は、SAGEによって予測されたものとよく一致した(表2)。タグP1およびP2は、それぞれアミラーゼおよびプレプロカルボキシペプチダーゼA2に対応することがわかった。GenBank R87には、プレプロカルボキシペプチダーゼA2に対するエントリーが全く存在せず、アミラーゼに対しては末端の切断された(truncated)エントリーダけしか存在せず、これにより、それらの指定を受けていない特性表示を説明できる。タグP3は、GenBankに含まれる既知の機能の遺伝子のいずれとも一致しなかつたが、多数のESTと一致し、このことにより、タグP3が真の(bona fide)転写産物を表すことがさらに明らかになった。P4によって同定されたcDNAは有意な相同性を全く示さず、このことは、そのcDNAがこれまでに特性決定されていない臍臓転写産物を表すことを示唆した。

表2

指定を受けていないSAGEタグの特性決定

<u>タグ</u>	<u>SAGE存在量</u>	<u>13mer Hyb</u>	<u>SAGEタグ</u>	<u>記載</u>
P1 TCCTCAAAA (配列番号 38)	1.7%	1.5% (6/388)	+	脾臓アミラーゼの3'末端 (M28443)
P2 TGGGAGACC (配列番号 39)	1.1%	1.2% (43/3700)	+	プロテオカルボキシペプチダーゼA2の 3'末端(U19977)
P3 ACGCAGGGGA (配列番号 40)	0.6%	0.2% (5/2772)	+	EST一致(R45808)
P4 AATTGAAGA (配列番号 41)	0.6%	0.4% (6/1587)	+	一致なし

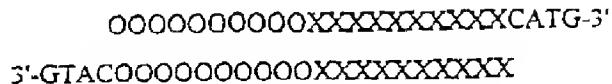
「タグ」および「SAGE存在量」は表1で説明したとおりであり、「13merHyb」は、上記のような13merでcDNAライブラリーをスクリーニングすることによって得られた結果を示す。スクリーニングした総ブラーク数で除した陽性ブラークの数を、存在量パーセントの後ろのカッコ内に示す。「SAGEタグ」の欄の+の符号は、発現されたSAGEタグ配列が、単離したクローンの3'末端付近で同定されたことを示す。「記載」とは、1995年6月9日のNCBIで毎日更新されているGenBankエントリーのBLAST検索の結果を示す(Altschul, ら, J.Mol. Biol. 215:403, 1990)。記載および受け入れ番号は最も顕著な一致をみたものについて示してある。P1は、末端の切断されたアミラーゼに対するエントリーに一致することがわかり、P2はプレプロカルボキシペプチダーゼA2に対する未発表のエントリー(これを、GenBank R87の後に参加させた)に一致することがわかつた。

実施例4

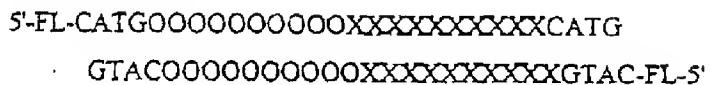
SAGEによって得られたジタグを、本明細書に記載されているように、PS

AまたはC Sにより分析できる。P S Aの1つの好ましい実施形態では、ジタグを用いて以下の工程を行う。

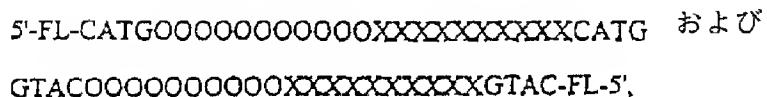
すなわち、先の実施例に記載したようにジタグを調製し、増幅し、アンカリング酵素で切断する。



突出部分に相補的な、確認物質（例えば、蛍光成分、FL）を含有する4塩基のオリゴマー（例えば、FL-CATG）を調製する。過剰のFL-CATGオリゴマーを上記のジタグに以下のように連結する。



次にそれらのジタグを精製し溶融して、例えば、式：



を有する一本鎖D N Aを得る。この一本鎖D N Aの混合物は好ましくは連続希釈する。各連続希釈物を、適切なストリンジエントな条件下で、グリッドに分注した一本鎖オリゴを含有する固相マトリックスとハイブリダイズさせる。それらのオリゴヌクレオチドの全部がアンカリング酵素切断配列の半分の部位を含有している。ここで用いた例では、オリゴヌクレオチド配列は5'末端にCATG配列を含有する。

CATG0000000000、CATGXXXXXXXX、等

（または、3'末端にCATGを含有する：0000000000CATG）

マトリックスは当業界で公知の如何なる物質から構成されていてもよく、オリゴヌクレオチド担持チップは当業界で公知の手順によって作製することができ、例えば、VLSIP法（Fodor ら、前掲）によって作製されたオリゴヌクレオチドを含有するシリコンチップが挙げられる。

オリゴヌクレオチド担持マトリックスは、グリッド内の各位置における蛍光

ジタグの存在または不存在について評価する。

1つの好ましい実施態様では、グリッドには 4^{10} または 1,048,576 の一般配列 C ATG0000000000 のオリゴヌクレオチドが存在し、あらゆる可能性のある、10 塩基配列が CATG の 3' 側に表示されるようになっている(ここでは、CATG は、ジタグの 3' 末端にあるアンカリング酵素の半分の部位に相補的であるアンカリング酵素の半分の部位の一例として用いられている)。ヒト・ゲノムには 100,000~200,000 ほどの様々な発現遺伝子があると推定されるので、ヒト・ゲノムにおける発現遺伝子由来の cDNAにおいて見出される最も 3' 側のアンカリング酵素部位に隣接している可能な配列の全部を検出するのに十分なオリゴヌクレオチド配列がある。

さらに別の実施態様では、次の配列：

プライマー A - GGAGGCATG(X)₁₀(O)₁₀CATGGATCC - プライマー B、

プライマー A - CCTCGTAC(X)₁₀(O)₁₀GTACGTAGG - プライマー B

を含有する上記の配列を増幅し、タギング酵素、次にアンカリング酵素で切断して、次の構造：

(O)₁₀CATG-3'

を有するタグ相補体を生成させる。次に、このタグ相補体を、標識し、融解し、固相支持体上でオリゴヌクレオチドとハイブリダイズさせることが可能である。

様々なライブラリー(示差的スクリーニングプローブを表す)間で、様々な希釈率で、グリッド上での蛍光プロフィールを比較することにより、示差的発現を測定した。例えば、

ライブラリー A, 1:10 に希釈したジタグ

ライブラリー B, 1:10 に希釈したジタグ

	A	B	C	D	E
1	FL				
2					FL
3		FL	FL		
4				FL	
5	FL				

	A	B	C	D	E
1	FL				
2				FL	
3	-	FL	FL		
4	-				
5	FL				FL

ライブラリーA, 1:50に希釈したジタグ

	A	B	C	D	E
1	FL				
2					
3		FL			
4				FL	
5	FL				

ライブラリーA, 1:100に希釈したジタグ

	A	B	C	D	E
1	FL	-			
2	-				
3		FL			
4				FL	
5	FL				

ライブラリーB, 1:50に希釈したジタグ

	A	B	C	D	E
1	FL				
2			FL		
3		FL	FL		
4					
5					

ライブラリーB, 1:100に希釈したジタグ

	A	B	C	D	E
1	FL				
2			FL		
3		FL			
4					
5				"	

かくして、個々のオリゴヌクレオチドは以下の特性を有するジタグにハイブリダイズする。

表3

希釈率	1:10		1:50		1:100	
	Lib A	Lib B	Lib A	Lib B	Lib A	Lib B
1A	-	+	-	+	+	+
2C		+		+		+
2E	+	+				
3B	-	+	-	+	-	-
3C	+	-		+		
4D	+		+		+	
5A	-	+	-		-	
5E		+				

表3は、示差的ハイブリダイゼーションの結果をまとめたものである。1A

および3Bにハイブリダイズしたタグは、示差的には発現されない、高い存在量のmRNAを反映し（その理由は、これらのタグが、全ての希釈率において双方

のライブラリーにハイブリダイズするからである) ; タグ2Cは、高い存在量のmRNA(しかし、ライブラリーBのみ)を同定する。2Eは、示差的に発現されることがわかっていない、低い存在量の転写産物を反映し(その理由は、それが最低の希釈率でのみ検出されるからである) ; 3Cは、ライブラリーBでの中程度の存在量の転写産物を反映し(その理由は、それが低いほうの2つの希釈率で発現されるからである)、この転写産物はライブラリーAでは低い存在量で発現される。4Dは、ライブラリーAに限って示差的に発現される高い存在量の転写産物を反映し; 5Aは、ライブラリーAでは高い存在量で発現されるが、ライブラリーBでは低い存在量でしか発現されない転写産物を反映し; 5Eは、ライブラリーBでのみ検出可能な、示差的に発現される転写産物を反映する。

別のP S Aの実施態様では、上記の工程3で蛍光性または他の確認物を使用する必要はないが、その代わりに、ジタグの最後の増幅のラウンドにおいて、標識d NTPを用いて、融解後に全分子の半分が標識されて、チップ上に固定されたオリゴヌクレオチドにハイブリダイズさせるためのプローブとして役立つようにする。

さらに別のP S Aの実施態様では、ジタグの代わりに、転写産物の特定の部分(例えば、転写産物の3'末端と第1アンカリング酵素部位との間の配列)を用いる。そのような特定の場合、「詳細な説明」で記載したように、二本鎖cDNA逆転写産物が作製される。この転写産物はアンカリング酵素で切断し、PCRプライマーを含有するリンカーを付加して、(このプライマーを一方の末端で用い、ポリA尾部を他方の末端で用いて)増幅を開始するが、この転写産物は依然としてストレプトアビジン・ビーズの上に存在したままである。増幅の最後のラウンドで、フルオレセイン化d NTPを用いて、分子の半分が標識されるようになる。この時点で、場合によっては、断片のサイズを小さくするために、リンカープライマーをアンカリング酵素を用いて除去してもよい。次に、先の実施例のように、これらの可溶性の断片を融解し、CATG0000000000を含有する固相マ

トリックスの上で捕捉する。(フルオレセイン化塩基を含有する断片の半分のみの)分析および評価は上記と同様である。

クローンの配列決定で使用するためには、（細胞クローニングのための限界希釈法において行われるように）ジタグまたはコンカテマーを希釈し、例えばマルチウェルプレートのウエルまたは他の受け器に添加して、ウエルが統計学上平均してウエル当たり1個未満のDNA分子を含有するようになる。次に、各ウエルには、PCRまたは他の增幅プロセス用の試薬を加え、各受け器中のDNAを例えばマススペクトルスコピーによって配列決定する。その結果は、単一の配列（その受け器中に単一の配列が存在していた）か、「ヌル」配列（DNAは存在していないなかった）であるか、あるいは二重配列（1より多いDNA分子）のいずれかであり、これは、二重配列はデータ分析の際には考慮から外されるであろう。その後、示差的発現の評価を、ここに記載したとおりに行う。

これらの結果は、SAGEが遺伝子発現についての定量的および定性的の両方のデータを提供することを実証するものである。種々の認識エレメントをもつ異なるアンカリング酵素および／またはタギング酵素を使用することにより、この戦略が大きな融通性を伴うようになる。特に、異なるアンカリング酵素は異なる部位でcDNAを切断するので、同じcDNA調製物の個々のサンプルに対して少なくとも2つの異なるアンカリング酵素を用いることにより、結果の確認が可能となり、またそれら酵素の1つに対する認識部位を含有しない可能性がある配列の分析の確認が可能となる。

完成に近づいているゲノムを十分に特性決定するための試みとして、SAGEは、あらゆる所与の細胞型または組織における発現の直接的な解読を可能にしなければならない。当座の間、SAGEの主な用途は、組織間での、および所与の細胞または組織の様々な発生段階および疾病状態での遺伝子発現パターンの比較であろう。PCRおよび手操作による配列決定を行う能力のある当業者であれば、この目的のためにSAGEを実施することが可能であろう。この技術を自動化配列決定装置に適用することにより、3時間の運転を1回行うだけで1,000個以上の転写産物を分析することが可能となろう。ABI377型配列決定装置では、

3時間の運転で36の鋳型について451bpの解読が得られる(451bp/11bp/タグ×3
6=1476タグ)。測定すべきタグの適切な数は、その応用例に依存する。例えば、

1つの組織では比較的高レベル（0.5%以上）で発現されるが他の組織では低いレベルで発現される遺伝子の決定には、たった1日あればよいであろう。細胞当たり100以上のmRNA（0.025%以上）で発現される転写産物の測定は、1つの測定装置で2、3ヶ月以内に定量可能でなければならない。2種の異なるアンカーリング酵素を用いることにより、所望の存在量の転写産物のほとんど全てが確実に同定されるであろう。それらの示差的提示(differential representation)に基づいて最も興味深いものであることがわかっているタグをコードする遺伝子は、表2で実証されているように、データベース検索とハイブリダイゼーションと配列分析とを組合せることによってポジティブに同定できる。明らかに、SAGEはヒト以外の生物の分析にも応用可能であり、特定の生物学的状態で発現される遺伝子の方に研究を導くであろう。

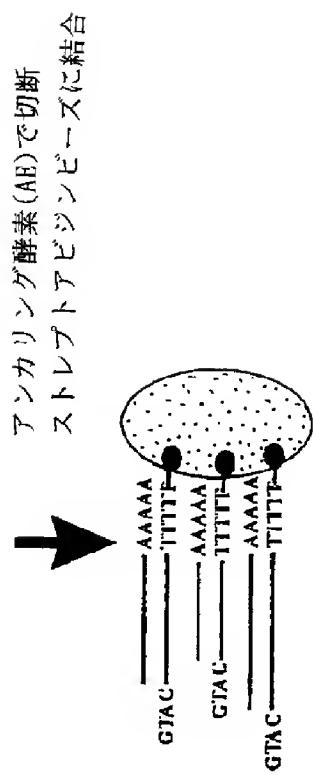
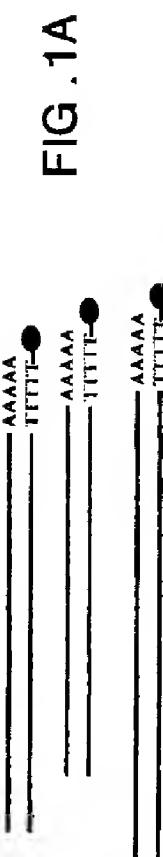
SAGEは、本明細書で記載したように、組織間での、または同じ組織の異なる段階間での、または病的な組織とその正常なものとの間での多数の遺伝子の発現の比較を可能にする。そのような分析は、例えば治療上、診断上および予後上関連のある遺伝子を同定するのに有用である。SAGE技術の多くの有用性の中でも、とりわけ治療上有用でありうる適切なアンチセンス試薬または3重らせん試薬の同定がある。さらに、遺伝子治療の候補もこのSAGE技術によって同定することが可能である。他の用途としては、遺伝子の発現が例えば疾病に対する素質、疾病の存在および疾病の予後と相関性があることがわかっている個々の遺伝子または遺伝子群を同定するための診断的使用が挙げられる。表1に示されるような存在量のプロフィールは、上記の用途に有用である。SAGEは、宿主内の生物（例えば、病原体）の検出、または宿主内で病原体によって発現される感染特異的遺伝子の検出にも有用である。

本発明におけるSAGEによって説明されるように、多数の発現遺伝子を短時間で同定できるということは、無限の用途を提供する。

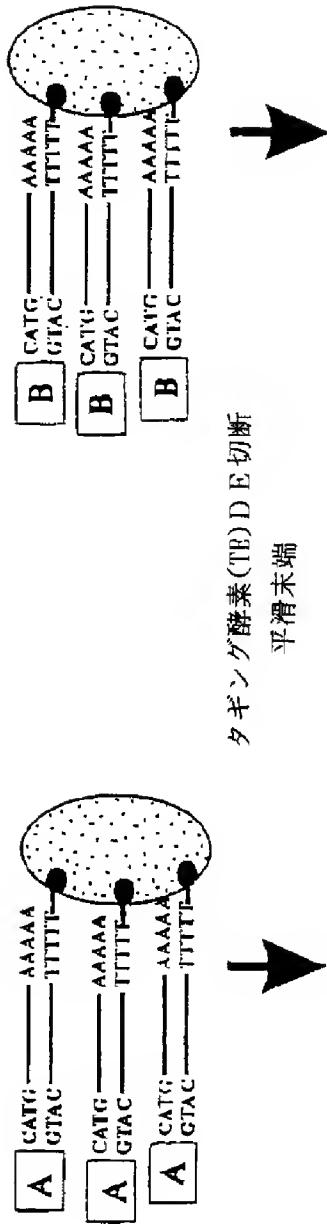
本発明は、現在好ましいとされる実施態様について記載してきたが、本発明

の精神を逸脱することなしに種々の改変が可能であると理解すべきである。したがって、本発明は、以下の特許請求の範囲によってのみ限定される。

[図1]

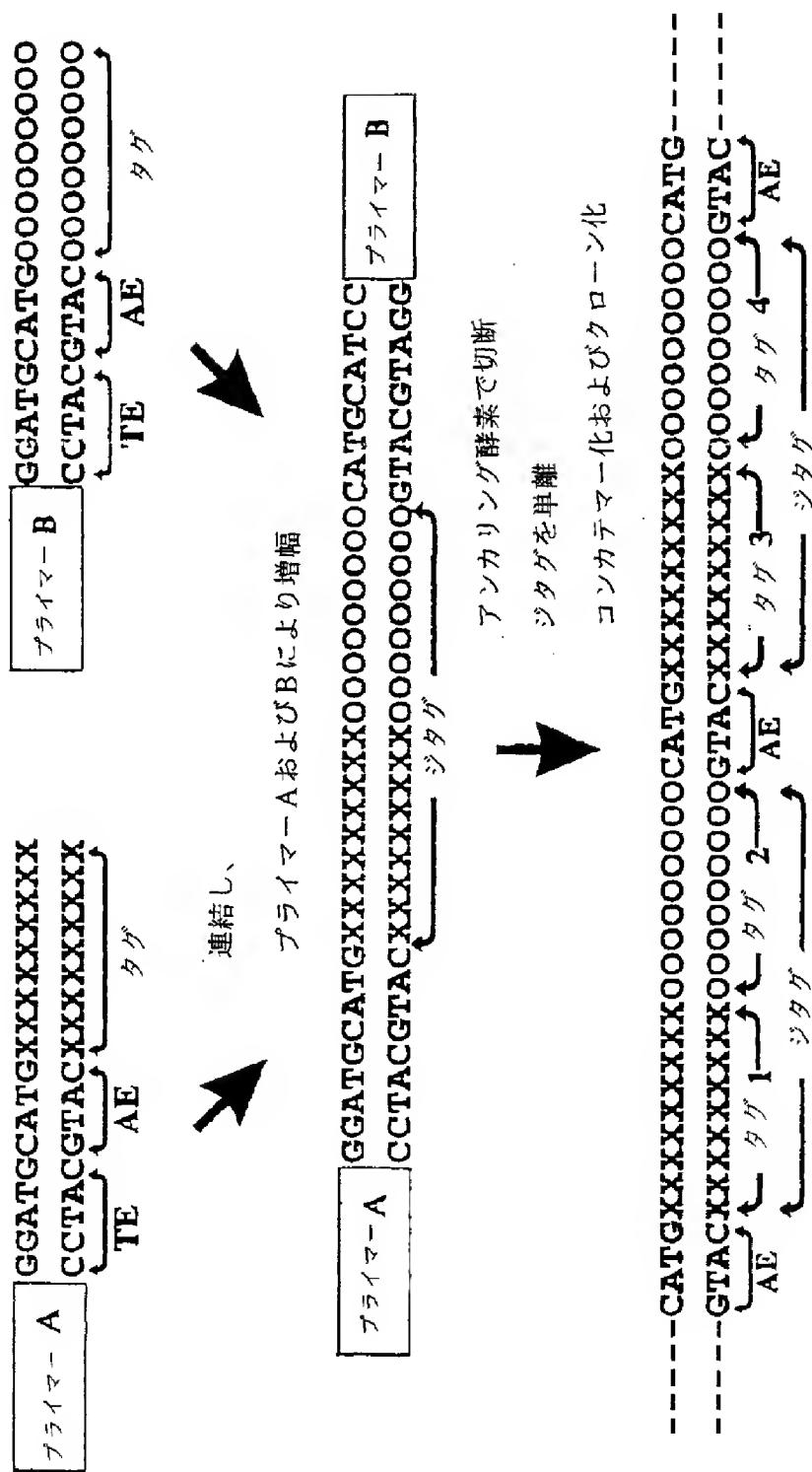


半分に分割
リンクマーク(A+B)に連結



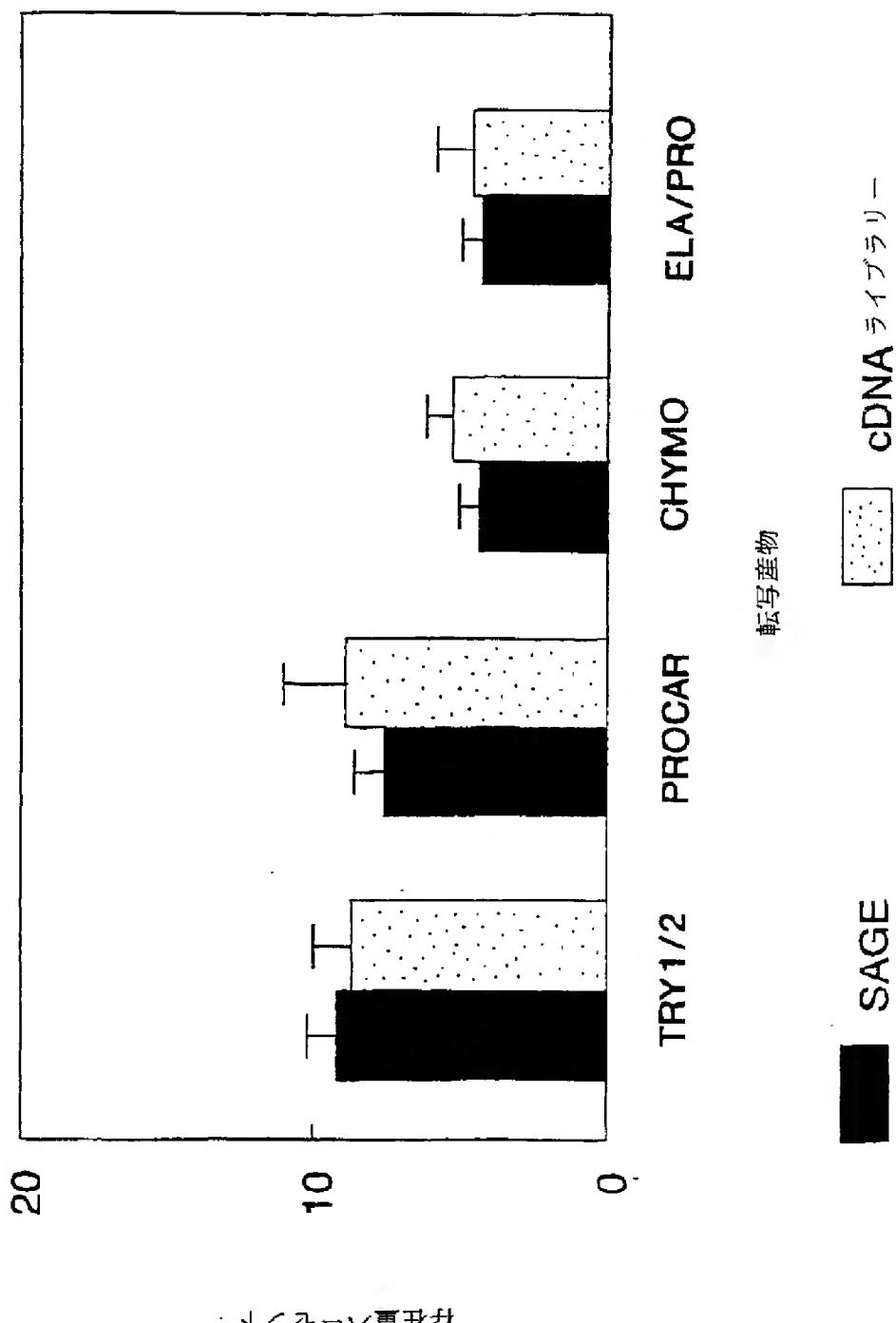
【図1】

FIG. 1B



【図2】

FIG.2



【図3】

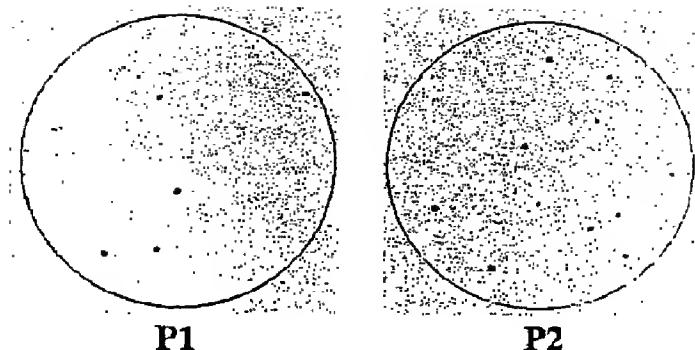


FIG. 3A

FIG. 3B

【図4】

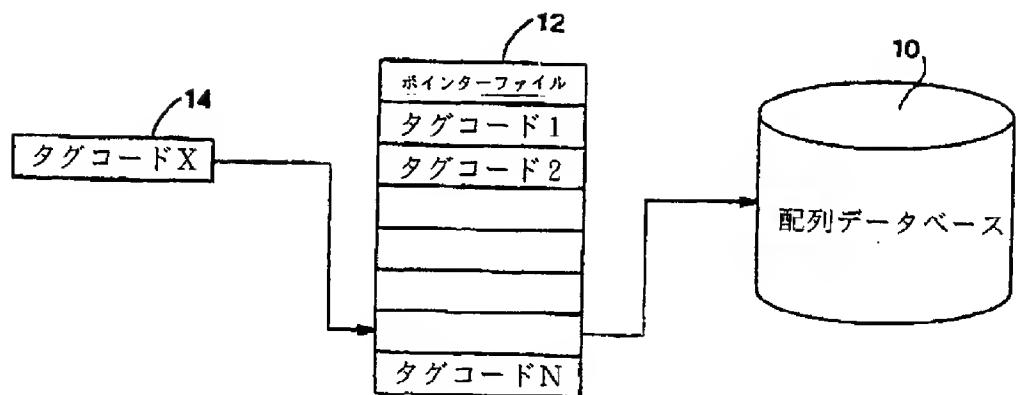
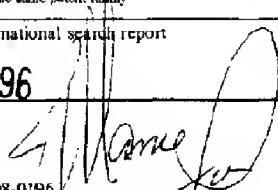


FIG. 4

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US96/14638
A. CLASSIFICATION OF SUBJECT MATTER		
IPC(6) :C12Q 1/68; C12P 19/34; C12N 15/00; C07H 21/00 US CL :Please See Extra Sheet.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/6, 91.1, 172.3; 536/22.1, 24.2, 24.3; 935/1, 5, 6, 8, 9, 16, 19		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS, DIALOG		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	US 5,508,169 A (K.V. DEUGAU) 16 April 1996, page 1, abstract.	1,4,6,11,18
X	UNRAU et al. Non-cloning amplification of specific DNA fragments from whole genomic DNA digests using DNA 'indexers'. Gene. 1994, Vol. 145, pages 163-169, especially the abstract and introduction.	1,4,6,11,18
X	WHITE et al. Concatemer chain reaction: A Taq DNA polymerase-mediated mechanism for generating long tandemly repetitive DNA sequences. Anal. Biochem. 1991, Vol. 199, pages 184-190, especially page 184, column 2, last paragraph, and page 186, column 1, last three lines.	1
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
<ul style="list-style-type: none"> * Special categories of cited documents. *A* document defining the general state of the art which is not considered to be of particular relevance *B* earlier document published on or after the international filing date *C* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *D* document referring to an oral disclosure, i.e. exhibition or other means *E* document published prior to the international filing date but later than the priority date claimed 		<ul style="list-style-type: none"> * T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention * X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone * Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art * & * document member of the same patent family
Date of the actual completion of the international search	Date of mailing of the international search report	
15 OCTOBER 1996	13 NOV 1996	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer BONNIE WEISS  Telephone No. (703) 308-0196	

INTERNATIONAL SEARCH REPORT		International application No. PCT/US96/14638
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KATO. Description of the entire mRNA population by a 3' end cDNA fragment generated by class IIIS restriction enzymes. Nucleic Acids Res. 1995, Vol. 23, No. 18, pages 3685-3690, especially the abstract and page 3690, second full paragraph.	1-45
Y	EP 0 679 716 A1 (MATSUBARA, KENICHI) 11 February 1995, page 1, abstract.	4,10,11
Y	SAMBROOK et al. Alternative methods of cloning cDNA. Molecular Cloning, A Laboratory Manual, second edition. New York: Cold Spring Harbor Laboratory Press, 1989, Vol. 2, pages 8.27-8.29, see entire document.	4,10,11
Y	SAMBROOK et al. Ligation of phosphorylated linkers to blunted target fragments. Molecular Cloning, A Laboratory Manual, second edition. New York: Cold Spring Harbor Laboratory Press, 1989, Vol. 3, pages F.9-F.11, see entire document.	4,10,11

INTERNATIONAL SEARCH REPORT

International application No. PCT/US96/14638

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 91.1, 172.3; 536/22.1, 24.2, 24.3; 935/1, 5, 6, 8, 9, 16, 19

フロントページの続き

(81)指定国 EP(AT, BE, CH, DE,
DK, ES, FI, FR, GB, GR, IE, IT, L
U, MC, NL, PT, SE), OA(BF, BJ, CF
, CG, CI, CM, GA, GN, ML, MR, NE,
SN, TD, TG), AP(KE, LS, MW, SD, S
Z, UG), UA(AM, AZ, BY, KG, KZ, MD
, RU, TJ, TM), AL, AM, AT, AU, AZ
, BB, BG, BR, BY, CA, CH, CN, CZ,
DE, DK, EE, ES, FI, GB, GE, HU, I
L, IS, JP, KE, KG, KP, KR, KZ, LK
, LR, LS, LT, LU, LV, MD, MG, MK,
MN, MW, MX, NO, NZ, PL, PT, RO, R
U, SD, SE, SG, SI, SK, TJ, TM, TR
, TT, UA, UG, UZ, VN

(72)発明者 ヴェルヴレスク, ヴィクター, イー.
アメリカ合衆国 21202 メリーランド州
バルチモア, エヌ. カルパート スト
リート 650, アパートメント シー

(72)発明者 ツアン, リン
アメリカ合衆国 21218 メリーランド州
バルチモア, セント ポール ストリー
ト 3100, アパートメント 405